Genome-wide DNA sampling by Ago nuclease from the cyanobacterium *Synechococcus elongatus*

Anna Olina, Anton Kuzmenko, Maria Ninova, Alexei A. Aravin, Andrey Kulbachinskiy, Daria Esyunina

Supplementary Information

Optimized nucleotide sequence of SeAgo

CAAGACCTGCTGTCTAACCTGCGGCGTAGCAGTATTGTCTTAAACAGATTTTACGTAAAATCCCTTTCTCAGT
CCGATCTTACCGCCTACGAGTACCGTTGTATCTTCAAAAAGACACCGGAGTTAGGCGATGAAAAGAGATTGTT
GGCTTCTATCTGCTATAAGTTAGGAGCGATCGCAGTTAGAATTGGCTCAAACATAATTACGAAGGAGGCCGTT
CGGCCTGAAAAATTGCAGGGTCATGACTGGCAATTAGTACAGATGGGAACCAAACAGTTAGATTGCCGGAATG
ACGCTCATAGATGCGCACTGGAGACGTTTGAAAGAAAATTCTTGGAACGCGACCTTTCGGCTTCGAGTCAAAC
CGAGGTCCGGAAAGCAGCTGAGGGGGGCCTTATATGGTGGGTAGTCGGGGCGAAGGGCATTGAAAAATCTGGC
AACGGTTGGGAAGTACATAGAGGCAGACGGATAGATGTCAGTTTGGACGCAGAAGGGAATTTGTACCTGGAGA
TAGACATTCATCATCGGTTCTATACACCGTGGACCGTTCACCAGTGGCTTGAACAGTACCCCGAAATTCCGTT
GTCCTACGTTCGGAACAACTATTTGGACGAAAGACATGGATTTATTAACTGGCAGTACGGTCGGTTCACTCAA
GAGCGGCCTCAGGACATATTGTTGGATTGTTTAGGGATGAGTTTAGCGGAATATCATCTGAATAAGGGTGCAA
CGGAGGAAGAAGTGCAGCAGTCTTATGTTGTCTATGTTAAGCCTATAAGTTGGAGAAAAGGTAAATTGACCGC
GCATTTGAGTCGCCGGCTTTCCCCATCTCTTACGATGGAGATGCTGGCCAAAGTCGCTGAGGATTCAACCGTA
TGTGACCGCGAAAAGCGTGAGATACGTGCGGTGTTTAAGTCGATAAAACAAAGTATAAATCAACGTTTGCAAG
AGGCTCAAAAGACCGCATCTTGGATATTGACCAAGACATACGGGATTTCGAGCCCAGCAATAGCTCTTTCTTG
CGATGGTTACCTTTTGCCGGCAGCGAAGCTGTTGGCAGCTAACAAACAGCCCGTGTCCAAGACGGCTGATATA
CGGAATAAAGGATGCGCGAAGATTGGGGAAACTAGCTTCGGGTACCTTAACTTATACAACAATCAACTGCAGT
ACCCCCTGGAAGTACATAAATGCTTGTTGGAGATCGCGAATAAGAACAATCTTCAGTTGAGCTTGGATCAACG
GCGTGTACTTAGCGATTACCCACAGGATGATCTTGACCAACAGATGTTCTGGCAGACGTGGTCATCGCAAGGC
ATAAAGACAGTCTTAGTTGTCATGCCGTGGGACTCACACCACGACAAGCAGAAGATCCGCATTCAGGCGATTC
AAGCTGGGATAGCGACCCAGTTTATGGTTCCGTTGCCTAAAGCAGACAAGTATAAGGCGTTAAATGTGACTCT
TGGGCTTCTTTGTAAAGCAGGATGGCAGCCAATCCAGTTGGAGAGCGTCGACCATCCTGAAGTCGCGGATTTG
ATCATAGGGTTCGACACGGGGACAAATCGTGAGTTATACTACGGTACCTCCGCATTTGCCGTCCTTGCGGATG
GCCAGTCCTTGGGTTGGGAATTACCTGCGGTCCAACGTGGAGAGACATTCAGCGGGCAAGCCATATGGCAAAC
CGTTAGTAAGCTTATAATTAAGTTTTACCAGATATGCCAACGCTATCCCCAGAAGCTGCTGTTAATGAGAGAT
GGACTTGTCCAAGAGGGGGAGTTTCAGCAAACGATCGAGCTTCTTAAAGAGCGCAAAATTGCAGTGGACGTCA
TATCGGTTAGAAAGAGTGGAGCAGGACGCATGGGCCAAGAAATCTACGAGAACGGGCAGTTAGTCTATCGGGA
TGCGGCCATTGGATCCGTTATTCTTCAACCGGCGGAGAGATCTTTCATTATGGTAACCTCGCAGCCAGTCTCG
AAGACTATCGGATCCATTCGTCCATTACGGATTGTTCACGAATACGGCAGTACTGACCTGGAGTTGTTGGCCT
TACAGACTTACCATCTTACGCAGCTTCATCCGGCAAGTGGATTCCGGTCGTGTCGCCTGCCGTGGGTCCTTCA
TTTGGCCGACCGGTCCTCTAAGGAGTTTCAGCGGATAGGACAGATATCCGTTCTTCAGAATATAAGTAGAGAC
AAACTGATCGCCGTATAA

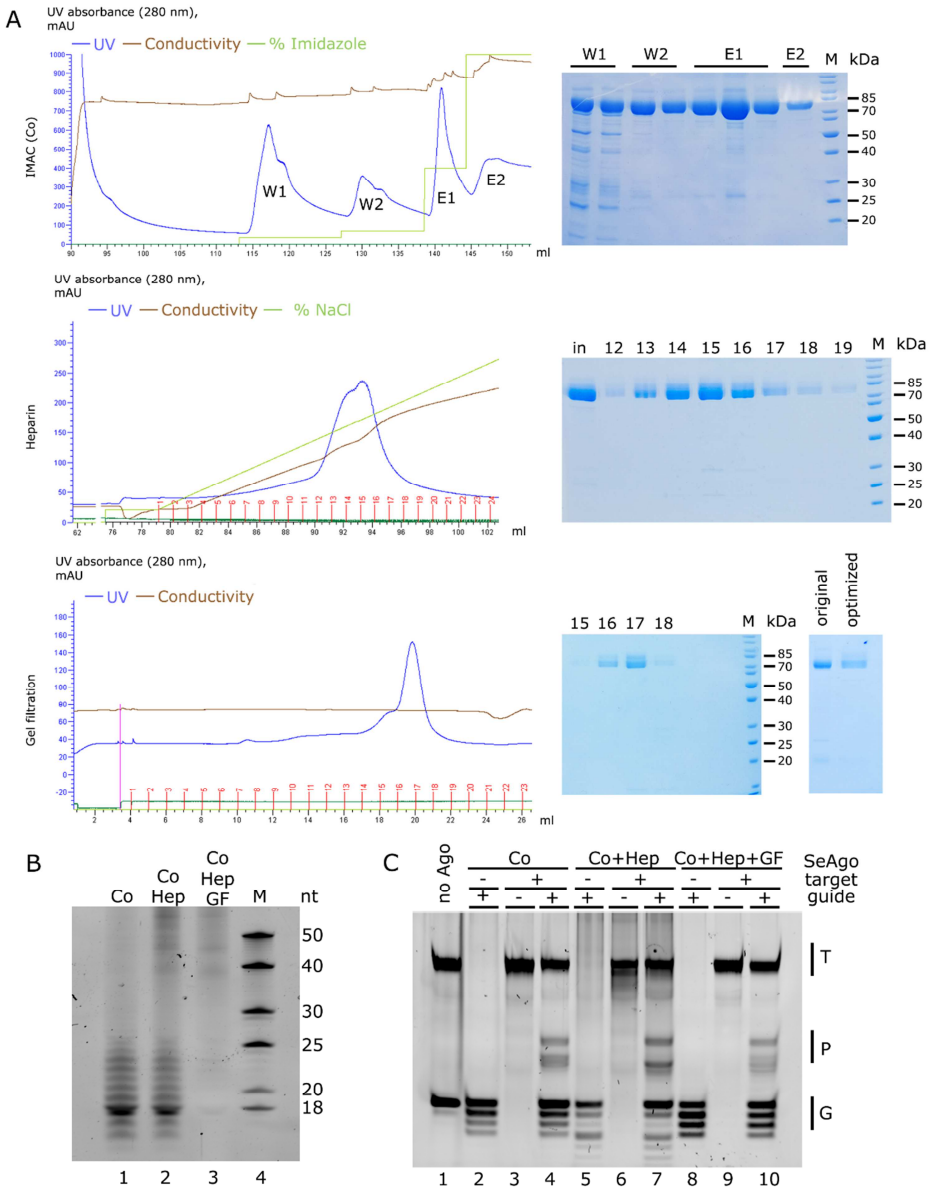

Figure S1. Purification of SeAgo. (A) Purification steps of the optimized SeAgo protein with corresponding chromatograms and Coomassie-stained protein gels (fraction numbers are indicated). *E. coli* cells expressing SeAgo were disrupted using French-press, the lysate was clarified by centrifugation and SeAgo was purified in 3 consecutive chromatography steps: metal-affinity chromatography on a Co^{2+}-crude column, affinity chromatography on a heparin column and size-exclusion chromatography on a Superdex 200 10/300 GL column (GE Healthcare). The right bottom gel shows comparison of purified SeAgo samples obtained after expression of the wild-type or codon-optimized genes. (B) SeAgo loses the majority of associated nucleic acids after purification. Equal amounts of protein after each purification step were subjected to phenol-chloroform extraction of nucleic acids, which were visualized by 15% denaturating PAGE stained with SYBR-gold. Most associated 15-25 nt small DNAs are lost after the 3rd step of purification. (C) The activity of SeAgo after each step of purification. Equal amounts of SeAgo purified in 1 (Co^{2+}-chromatography), 2 (Co^{2+} and heparin) or 3 steps (Co^{2+}, heparin and gel filtration) were incubated with guide ('G', T-guide) and/or target ('T', T-tDNA_Cy3) DNAs. Positions of the reaction products ('P') are indicated.

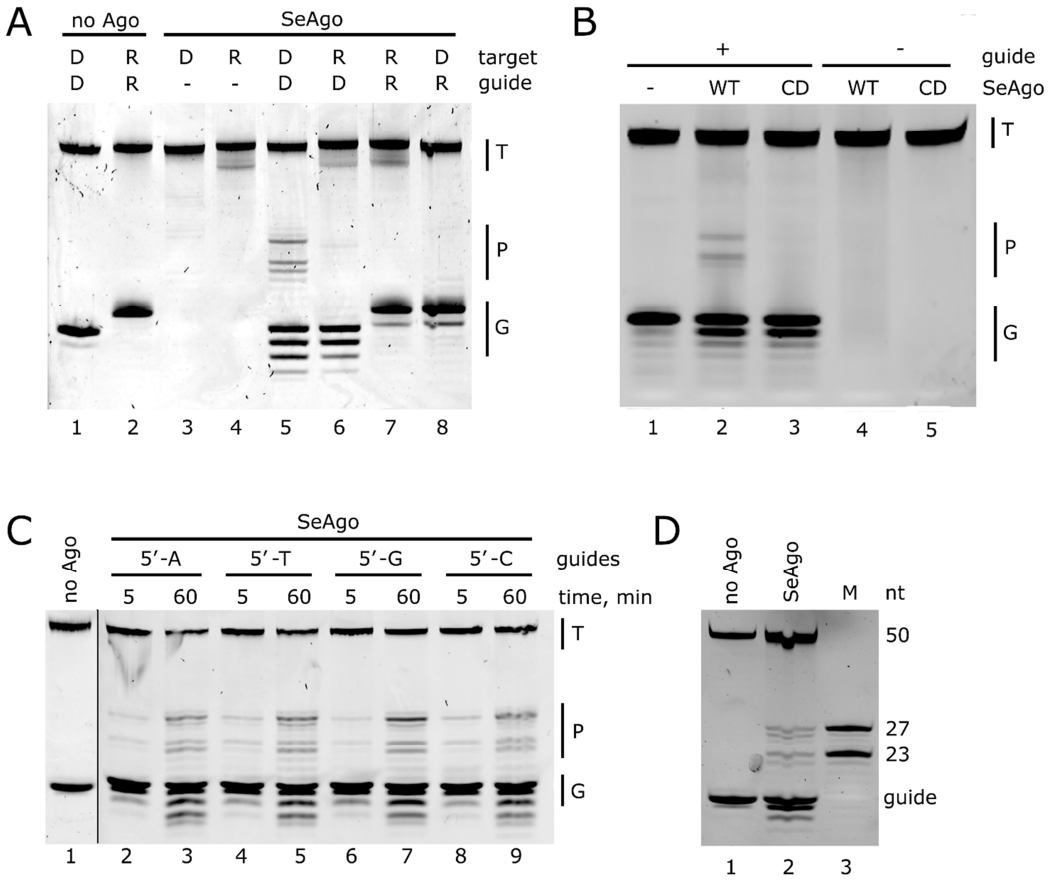

Figure S2. Analysis of the catalytic properties of SeAgo. (A) SeAgo is a DNA-guided DNA endonuclease. The cleavage assays were performed with various guide and target combinations, DNA (‘D’, T-tDNA_Cy3) or RNA (‘R’), for 60 min at 37°C. (B) Comparison of the activities of wild-type (WT) and catalytically dead (CD) SeAgo variants. The reaction was carried out for 60 min at 37 °C. (C) The 5’-nucleotide in guide DNA does not affect the SeAgo activity. Analysis of DNA cleavage was performed for 5 and 60 min at 37 °C with guide variants containing different 5’-nucleotides and corresponding complementary target DNAs (T-tDNA, A-tDNA, C-tDNA and G-tDNA, see Table S1). (D) Mapping of the cleavage site with markers of specific lengths. The reaction was performed with the G-tDNA target and G-guide oligonucleotides (Table S1), producing the 23 nt 5’- and 27 nt 3’-fragments of the target. 23 nt and 27 nt oligonucleotides of corresponding sequences were applied as size markers. Positions of the target (‘T’), guide (‘G’) and product (‘P’) DNAs are indicated.

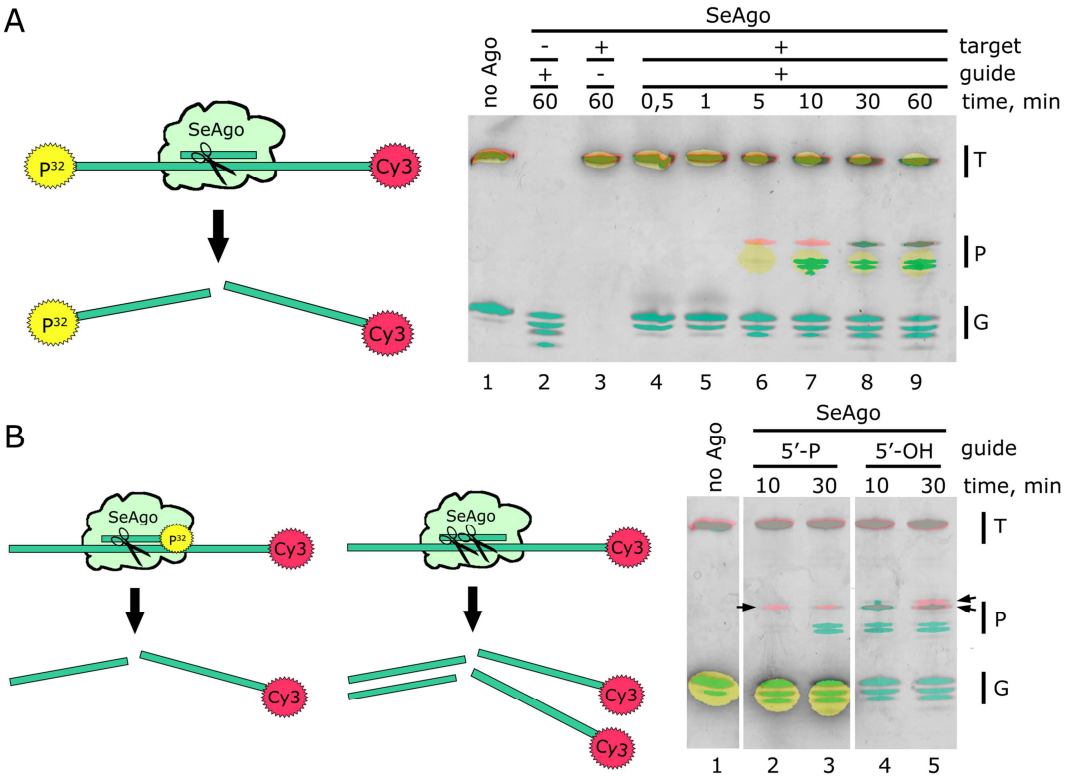

Figure S3. Analysis of the SeAgo activity with differentially labeled guide and target DNAs. (A) The target DNA containing a 3'-Cy3 fluorophore (T-tDNA_Cy3, Table S1) was 5'-labeled using γ-^{32}P-ATP and polynucleotide kinase. The T-guide DNA was 5'-phosphorylated with nonradioactive ATP. Guide-loaded SeAgo was incubated with target DNA for different time intervals and the reaction products were analyzed by 15% denaturing PAGE. The gel was first scanned in the Cy3 channel (pink), then stained with SYBR-gold and scanned in corresponding channel (turquoise), and then placed in a ^{32}P-sensitive screen followed by phosphorimaging with a Typhoon FLA 9500 scanner (yellow). The overlay of the three images is shown. After initial target cleavage by SeAgo, the 25 nt 5'-product DNA undergoes subsequent 1-2 nt cleavage from the 3'-end. The 25 nt 3'-product migrates higher because of the presence of the 3'-Cy3 group and remains stable during the course of the reaction. (B) Comparison of the target DNA cleavage (T-tDNA_Cy3) with 5'-phosphorylated (lanes 1-3, labeled with γ-^{32}P-ATP) and 5'-OH (lanes 4-5) T-guide variants. In the case of the non-phosphorylated guide variant, an additional 26 nt Cy3-labeled cleavage product appeared after 30 min incubation (pink). Thus, in the absence of the 5'-phosphate group in guide DNA the cut site can be shifted by 1 nt closer to the 5'-end of the target.

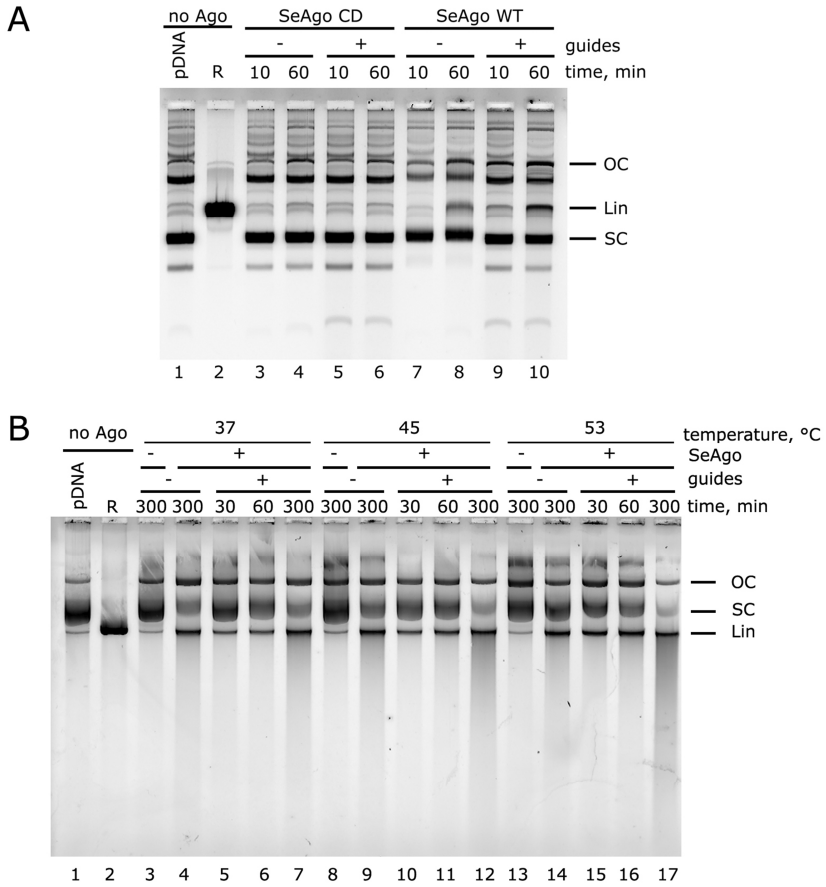

Figure S4. Activity of SeAgo on plasmid DNA. (A) Plasmid linearization depends on the catalytic activity of SeAgo. The pJET_target plasmid was incubated with wild-type (WT) and catalytically-dead (CD) SeAgo variants in the absence or in the presence of a pair of guide DNAs corresponding to the target region (Fig. S2, Table S1). The reaction was carried out for indicated time intervals at 37 °C. The wild-type but not catalytically dead SeAgo linearizes plasmid DNA after 1 hour incubation. (B) The activity of the SeAgo protein on the pSRKKm plasmid was measured for indicated time intervals at different temperatures in the absence or in the presence of 2 pairs of guide DNAs, separated by 1 kb, as described previously (1). The supercoiled plasmid undergoes linearization with further processing and appearance of a smear of shorter DNA fragments, depending on the presence of SeAgo. This activity is largely independent of the guide molecules. Positions of the supercoiled (SC), relaxed (OC) and linearized (Lin) plasmid substrates are shown on the right of the gel (note that in the case of pSRKKm the positions of the linear and supercoiled forms are reversed on the gel, see (1)). R, control linearized plasmids obtained by treatment with a restriction endonuclease.

1. Kuzmenko, A., Yudin, D., Ryazansky, S., Kulbachinskiy, A. and Aravin, A.A. (2019) Programmable DNA cleavage by Ago nucleases from mesophilic bacteria Clostridium butyricum and Limnothrix rosea. *Nucleic acids research*, **47**, 5822-5836.

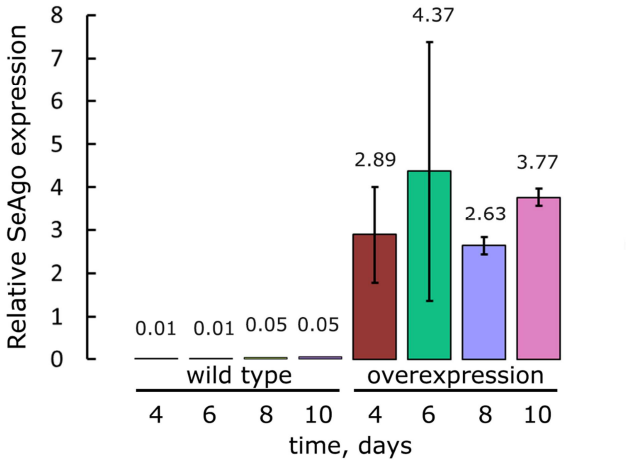

Figure S5. RT-qPCR analysis of the SeAgo expression in the wild-type and overexpression *S. elongatus* strains. The overexpression strain contained an additional copy of the SeAgo gene placed in the NS1 site under control of the PSBAII promoter (Fig. 3A). The amount of SeAgo mRNA was normalized to the expression of the housekeeping *rnpB* gene. Averages from two-three independent measurements are shown. The expression of SeAgo in both strains is relatively stable at different time points. The level of SeAgo mRNA in the overexpression strain is ~200 times higher than in wild-type cells and is comparable to the expression of the reference housekeeping gene.

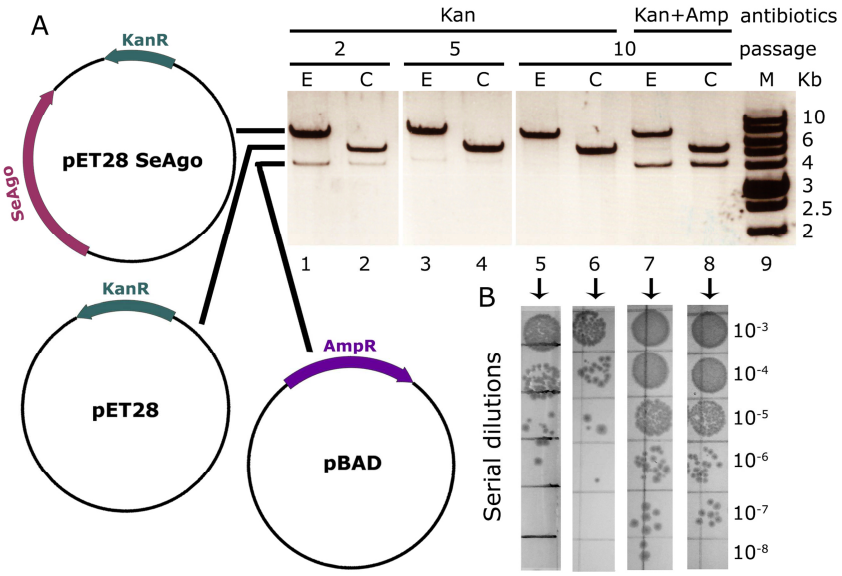

Figure S6. Analysis of the effects of SeAgo on plasmid maintenance in *E. coli*. The experiment was performed with codon-optimized SeAgo gene cloned into the pET28 expression vector under the control of the T7 RNAP promoter. *E. coli* BL21(DE3) cells were co-transformed with pET28_SeAgo (Kan^R) and pBAD (Amp^R) plasmids (experiment, E) or with empty pET28 and pBAD (control, C) and cultivated in the LB medium for several passages. During each passage, the cells were grown in liquid LB with 0.2 mM IPTG and kanamycin at 20°C for 24 hours and then transferred to fresh LB with IPTG and kanamycin with a 1:100 dilution. The cultivation was performed either in the absence or in the presence of ampicillin, either allowing or preventing the loss of the pBAD vector. An aliquot of cell culture was collected by centrifugation after each passage. Part of the aliquots was used for plasmid DNA purification and another part was mixed with glycerol and frozen in liquid nitrogen. (A) Plasmid DNA was linearized by the NcoI restriction endonuclease and analysed by 1% agarose gel electrophoresis. Positions of linearized plasmids are indicated. (B) At the end of the experiment, the frozen cells were thawed on ice, serially diluted with fresh LB medium (up to 10^{-11}) and plated on agar LB dishes with ampicillin and kanamycin (shown on the figure) or kanamycin only (as a control for equal cell densities). The results demonstrate that the pBAD plasmid is similarly lost from the cells independently of the presence of the SeAgo gene in pET28 (compare samples 5 and 6 in A and B). In both cases, the difference in the pBAD content between the cells cultivated in the absence and in the presence of ampicillin is about 3 orders of magnitude after the 10^{th} passage (compare samples 5/6 and 7/8).

Table S1. Sequences of oligonucleotides used in the cleavage assays. The target DNA nucleotides surrounding the cleavage position are bold underlined.

Oligonucleotide name	Sequence	Description
1. G-guide	5’-GTTAGACTTTAAGTCAAT	guide forms 5'-G pair with G-tDNA
2. C-guide	5’-CTTAGACTTTAAGTCAAT	guide forms 5'-C pair with C-tDNA
3. A-guide	5’-ATTAGACTTTAAGTCAAT	guide forms 5'-A pair with A-tDNA
4. T-guide	5’-TTTAGACTTTAAGTCAAT	guide forms 5'-T pair with T-tDNA
5. g38NT_mm1	5’-CTTAGACTTTAAGTCAAT	guide forms mismatched pair in position 1 with G-tDNA
6. g38NT_mm2	5’-GATAGACTTTAAGTCAAT	guide forms mismatched pair in position 2 with G-tDNA
7. g38NT_mm3	5’-GTAAGACTTTAAGTCAAT	guide forms mismatched pair in position 3 with G-tDNA
8. g38NT_mm4	5’-GTTTGACTTTAAGTCAAT	guide forms mismatched pair in position 4 with G-tDNA
9. g38NT_mm5	5’-GTTACACTTTAAGTCAAT	guide forms mismatched pair in position 5 with G-tDNA
10. g38NT_mm6	5’-GTTAGTCTTTAAGTCAAT	guide forms mismatched pair in position 6 with G-tDNA
11. g38NT_mm7	5’-GTTAGAGTTTAAGTCAAT	guide forms mismatched pair in position 7 with G-tDNA
12. g38NT_mm8	5’-GTTAGACATTAAGTCAAT	guide forms mismatched pair in position 8 with G-tDNA
13. g38NT_mm9	5’-GTTAGACTATAAGTCAAT	guide forms mismatched pair in position 9 with G-tDNA
14. g38NT_mm10	5’-GTTAGACTTAAAGTCAAT	guide forms mismatched pair in position 10 with G-tDNA
15. g38NT_mm11	5’-GTTAGACTTTTAGTCAAT	guide forms mismatched pair in position 11 with G-tDNA
16. g38NT_mm12	5’-GTTAGACTTTATGTCAAT	guide forms mismatched pair in position 12 with G-tDNA
17. g38NT_mm13	5’-GTTAGACTTTAACTCAAT	guide forms mismatched pair in position 13 with G-tDNA
18. g38NT_mm14	5’-GTTAGACTTTAAGACAAT	guide forms mismatched pair in position 14 with G-tDNA
19. g38NT_mm15	5’-GTTAGACTTTAAGTGAAT	guide forms mismatched pair in position 15 with G-tDNA
20. g38NT_mm16	5’-GTTAGACTTTAAGTCTAT	guide forms mismatched pair in position 16 with G-tDNA
21. g38NT_mm17	5’-GTTAGACTTTAAGTCATT	guide forms mismatched pair in position 17 with G-tDNA
22. g38NT_mm18	5’-GTTAGACTTTAAGTCAAA	guide forms mismatched pair in position 18 with G-tDNA
23. G-tDNA	5’-	50 nt target DNA for G-guide

	TTTATCAAAAAGAGTATTGACT**TA**<u>AAGTCTAACCTATAGGATACTTAC</u>AG	
24. C-tDNA	5'-TTTATCAAAAAGAGTATTGACT**TA**<u>AAGTCTAAGCTATAGGATACTTAC</u>AG	50 nt target DNA for C-guide
25. A-tDNA	5'-TTTATCAAAAAGAGTATTGACT**TA**<u>AAGTCTAATCTATAGGATACTTAC</u>AG	50 nt target DNA for A-guide
26. T-tDNA	5'-TTTATCAAAAAGAGTATTGACT**TA**<u>AAGTCTAAACTATAGGATACTTAC</u>AG	50 nt target DNA for T-guide
27. T-tDNA_Cy3	5'-AATTTATCAAAAAGAGTATTGACT<u>**TA**AAGTCTAAACTATAGGATACTTAC</u>-Cy3	50 nt fluorescently labelled target DNA for T-guide
28. RNA guide	5'-GUUAGACUUUAAGUCAAU	18 nt RNA guide for the guide/target specificity assay
29. RNA target	5'-UUUAUCAAAAAGAGUAUUGACU<u>**U** **A**AAGUCUAACCUAUAGGAUACUU</u>ACAG	50 nt RNA target for the guide/target specificity assay
30. T guide plasmid	5'-GCCATCCCAATCGACAC	plasmid cleavage assay
31. NT guide plasmid	5'-GTGTCGATTGGGATGGC	plasmid cleavage assay
32. pJET-target insert	5'-CAGTGAATTCTATTTGGATCCAGA TCCCGAAAATTTATCAAAAAGAGT ATTGACTTAAAGTCTAACCTATAG GATACTTACAGCCATCGAGAGGGA CACGGCGAATA<u>GCCATCCCAATCG ACAC</u>CGGGGTCCGGGATCTGGATC TGGATCGCTAATAACAGGCCTGCT GGTAATCGCTCGACTAGTCACCTG AGTCGGTA	sequence of the insert in pJET_target, the site complementary to the T and NT gDNAs is underlined
33. qPCR_ago_d	5'-TGTTTTTAAGAGCATCAAACAGTC TATTAACCAGC	qPCR primers for the SynAgo gene
34. qPCR_ago_r	5'-AATATCCGCAGTCTTGGAGACAGG	
35. qPCR_rnpB_d	5'-CCTACGACGGAGGCTGCAAG	qPCR primers for the reference gene *rnpB*
36. qPCR_rnpB_r	5'-TGGATGATCTCCGCCCGAGT	

Table S2. Mass-spectrometry analysis of proteins co-precipitated with SeAgo from *S. elongatus*. The first 15 hits, sorted by the number of total independent spectra are shown.

No	Protein name	Description	Gene number	Mr, Da	Protein coverage, %	Peptides number	Protein probability	Total independent spectra	Initial probability	N instances
1	Uncharacterized protein	Single-stranded DNA binding	Synpcc7942_0079	16055	51.4	10	1	120	0.999	36
2	Single-stranded DNA-binding protein	Single-stranded DNA binding, DNA replication	Synpcc7942_0301	13556	56.7	10	1	93	0.999	21
3	Uncharacterized protein	Unreviewed	Synpcc7942_1632	18434	19.5	3	1	33	0.999	6
4	Elongator protein 3/MiaB/NifB	Cobalamin binding, iron-sulfur cluster binding, metal ion binding	Synpcc7942_1621	60064	2.9	1	0.9996	21	0.999	21
5	Uncharacterized protein	Unreviewed	Synpcc7942_0100	20270	8.2	2	1	7	0.9861	5
6	L-threonine synthase	Threonine synthase activity, lyase	Synpcc7942_1782	46721	12.6	2	1	5	0.999	1
7	Cellulose synthase	Cellulose synthase (UDP-forming) activity	Synpcc7942_2151	83495	2	1	0.9996	5	0.999	5
8	Response regulator receiver domain protein	Phosphorelay sensor kinase activity	Synpcc7942_1815	46477	4.1	1	0.9996	4	0.999	4
9	Protein serine/threonine phosphatase	Protein serine/threonine phosphatase activity	Synpcc7942_1515	32711	6.4	1	0.9996	4	0.999	4
10	Uncharacterized protein	Unreviewed	Synpcc7942_0605	37247	4.3	1	0.9996	4	0.999	4
11	Probable glycosyltransferase	Transferase activity	Synpcc7942_2018	40377	9.9	2	1	3	0.999	1
12	Iron-regulated ABC transporter ATPase subunit SufC	Oxidoreductase activity	Synpcc7942_2036	37144	10.7	2	1	3	0.999	2
13	Ribonuclease, Rne/Rng family	Ribonuclease activity, RNA binding	Synpcc7942_0878	87431	6.1	3	1	3	0.999	1
14	3-hydroxyacyl-[acyl-carrier-protein] dehydratase FabZ	Involved in unsaturated fatty acids biosynthesis	Synpcc7942_0930	16900	12.9	1	0.9996	3	0.999	3
15	Uncharacterized protein	GTP binding	Synpcc7942_1154	48181	3	1	0.9996	3	0.999	3